COMPARISON OF EXPERIMENTS

DAVID BLACKWELL
HOWARD UNIVERSITY

1. Summary

Bohnenblust, Shapley, and Sherman [2] have introduced a method of comparing two sampling procedures or experiments; essentially their concept is that one experiment α is more informative than a second experiment β, $\alpha \supset \beta$, if, for every possible risk function, any risk attainable with β is also attainable with α. If α is a sufficient statistic for a procedure equivalent to β, $\alpha > \beta$, it is shown that $\alpha \supset \beta$. In the case of dichotomies, the converse is proved. Whether $>$ and $\supset$ are equivalent in general is not known. Various properties of $>$ and $\supset$ are obtained, such as the following: if $\alpha > \beta$ and γ is independent of both, then the combination $(\alpha, \gamma) > (\beta, \gamma)$. An application to a problem in 2×2 tables is discussed.

2. Definitions

An *experiment* α is a set of N probability measures $u_1, \ldots, u_N$ on a Borel field B of subsets of a space X. The N measures are considered as N possible distributions over X, and performing the experiment consists of observing a sample point $x \in X$. A *decision problem* is a pair (α, A), where A is a bounded subset of N-space. The points $a \in A$ are considered as the possible actions open to the statistician; the loss from action $a = (a_1, \ldots, a_N)$ is a_i if the actual distribution of x is u_i. A *decision procedure* f for (α, A) is a B-measurable function from X into A, specifying the action a to be taken as a function of the sample point x obtained by the experiment. With every $f = [a_1(x), \ldots, a_N(x)]$ is associated a loss vector

$$v(f) = \left(\int a_1(x)\, du_1, \ldots, \int a_N(x)\, du_N\right);$$

the i-th component of $v(f)$ is the expected loss from f if x has distribution u_i. The range of $v(f)$ is a subset of N-space which we denote by $R_1(\alpha, A)$; the convex closure of $R_1(\alpha, A)$ will be denoted by $R(\alpha, A)$ and will be called the set of *attainable loss vectors* in (α, A); every vector in R is either attainable or approximable by a randomized mixture of $N + 1$ decision procedures.

THEOREM 1. $R(\alpha, A) = R(\alpha, A_1) = R_1(\alpha, A_1)$, *where* A_1 *is the convex closure* of A.

This theorem permits us to restrict attention to closed convex A, which we shall do in the following sections. The proof of the theorem will not be given here; it is straightforward except for the fact that $R(\alpha, A_1) = R_1(\alpha, A_1)$. This fact follows from the result that whenever A is closed, so is $R_1(\alpha, A)$, which has been proved elsewhere by the author [1].

Following Bohnenblust, Shapley and Sherman [2], we shall say that α is *more informative* than β, written $\alpha \supset \beta$, if for every A we have $R(\alpha, A) \supset R(\beta, A)$.

It is an immediate consequence of theorem 1 that if $R(\alpha, A) \supset R(\beta, A)$ for every closed convex A, then $\alpha \supset \beta$.

3. Conditions equivalent to $\alpha \supset \beta$

THEOREM 2. *The following conditions are equivalent to* $\alpha \supset \beta$.

(1) *For every* A *and every* $v \in R(\beta, A)$, *there is a* $v^* \in R(\alpha, A)$ *with* $v_i^* \leqq v_i$ *for all* i.

(2) *For every* A *and every choice of* $c_i \geqq 0$, $\sum c_i = 1$,

$$\min_{v \in R(\alpha, A)} \sum_i c_i v_i \leqq \min_{v \in R(\beta, A)} \sum_i c_i v_i .$$

(3) *For every* A,

$$\min_{v \in R(\alpha, A)} \sum_i v_i \leqq \min_{v \in R(\beta, A)} \sum_i v_i .$$

(4) *For every* A,

$$\min_{v \in R(\alpha, A)} (\max_i v_i) \leqq \min_{v \in R(\beta, A)} (\max_i v_i) .$$

PROOF. The implications $\alpha \supset \beta \rightarrow (1) \rightarrow (2) \rightarrow (3)$, $(1) \rightarrow (4)$ are immediate. We show that (3) implies $\alpha \supset \beta$. Let $d_1, \ldots, d_N$ be any constants, and let T be the linear transformation $Tv = (d_1 v_1, \ldots, d_N v_N)$. Then $R(\alpha, TA) = TR(\alpha, A)$ and $\min_{v \in R(\alpha, TA)} \sum v_i = \min_{v \in R(\alpha, A)} \sum d_i v_i$, and similarly for β. Thus (3) yields that for all A, $d_1, \ldots, d_N$, $\min_{v \in R(\alpha, A)} \sum_i d_i v_i \leqq \min_{v \in R(\beta, A)} \sum d_i v_i$: every supporting hyperplane of $R(\alpha, A)$ lies on one side of $R(\beta, A)$, so that $R(\alpha, A) \supset R(\beta, A)$. Finally, we show that (4) implies (2). For any A and any $c_i \geqq 0$, $\sum_i c_i = 1$, let $v_0 \in R(\beta, A)$ be a point where $\sum c_i v_i$ assumes its minimum value over $R(\beta, A)$, and let U be the linear transformation $Uv = v - v_0$. Then

$$\min_{v \in R(\beta, UA)} \sum_i c_i v_i = 0 = \min_{v \in R(\beta, UA)} (\max_i v_i) .$$

Applying (4) to UA yields $\min_{v \in R(\alpha, UA)} (\max_i v_i) \leqq 0$, so that $\min_{v \in R(\alpha, UA)} \sum_i c_i v_i \leqq 0$.

Thus $\min_{v \in R(\alpha, A)} \sum_i c_i v_i = \sum_i c_i v_{0i}$ so that (2) holds.

4. Reduction to standard experiment

For any α, let $p_i(x)$, $i = 1, \ldots, N$, be the density of u_i with respect to $Nu_0 = u_1 + \ldots + u_N$, so that for any $S \in B$, $u_i(S) = \int_S Np_i(x) du_0$. Then $p_i \geqq 0$, $\sum p_i = 1$ except on a set of u_0 measure zero, and we may redefine p_i here so that the conditions hold identically. Let P be the set of N-tuples $p = (p_1, \ldots, p_N)$, $p_i \geqq 0, \sum p_i = 1$, and define, for any Borel subset of A of P, $m_i(A) = u_i\{p(x) \in A\}$, where $p(x) = [p_1(x), \ldots, p_N(x)]$, so that m_i is the distribution of p when x has distribution u_i. Since $p(x)$ is a sufficient statistic for x, considering i as the parameter, we would expect that the experiment α^* with measures $m_1, \ldots, m_N$ on P is equivalent to α. This fact was noted in [2] for the case in which the set A of actions has only a finite number of extreme points, and is embodied in

THEOREM 3. *For every* A, $R(\alpha, A) = R(\alpha^*, A)$.

PROOF. We shall use the notation $f \in (\alpha, A)$ to indicate that f is a decision procedure for the experiment (α, A). For any $f^* = [a_1(p), \ldots, a_N(p)] \in (\alpha^*, A)$, define $f = \{a_1[p(x)], \ldots, a_N[p(x)]\}$, so that $f \in (\alpha, A)$. Since p has the same distribution on P with respect to m_i, $i = 0, \ldots, N$, $Nm_0 = m_1 + \ldots + m_N$, as $p(x)$ on X with respect to u_i, for any Borel function $g(p)$ we have $\int g(p)dm_i = \int g[p(x)]du_i$. Choosing $g(p) = a_i(p)$ yields $v(f^*) = v(f)$, so that $R(\alpha^*, A) \subset R(\alpha, A)$. For the reverse inclusion, let $f = [a_1(x), \ldots, a_N(x)] \in R(\alpha, A)$, let $a_i^*(p) = E(a_i | p)$, the conditional expectation of a_i given p, with u_0 as the basic probability measure on X, and let $f^* = [a_1^*(p), \ldots, a_N^*(p)]$. Then for any Borel function $g(p)$, we have $\int a_i(x)g(p)du_0 = \int a_i^*(p)g(p)dm_0$. Choosing $g(p) = p_i$ and using $\int a_i(x)p_i du_0 = \int a_i(x)du_i$ and $\int a_i^*(p)p_i dm_0 = \int a_i^*(p)dm_i$ yields that $v(f) = v(f^*)$; it remains to show that $f^* \in (\alpha^*, A)$, that is, that the values of f^* are in A. If not, there is a linear function $L(a)$ with $L(a) \leqq 0$ for $a \in A$, $u_0\{L[f^*(x)] > 0\} > 0$. Then $\int L[f(x)]du_0 \geqq 0$, while $\int_S L[f^*(p)]du_0 > 0$, where $S = \{L[f^*(x)] > 0\}$, so that the two integrals cannot be equal, contrary to the definition of conditional expectation. Thus $f \in (\alpha^*, A)$, and the proof is complete.

Thus every experiment α is equivalent in the sense of theorem 3 to be an experiment α^* whose outcome is a point $p \in P$. The experiment α^* is called the *standard experiment* associated with α. Note that the measures $m_1, \ldots, m_N$ of the standard experiment α^* are completely determined by $m_0 = (m_1 + \ldots + m_N)/N$, since the density of m_i with respect to Nm_0 is simply p_i, and that the standard experiment associated with α^* is simply α^*. Moreover, any probability measure m_0 over P such that $\int Np_i dm_0 = 1$ for $i = 1, \ldots, N$ is the m_0 of a standard experiment α^*, with $m_1, \ldots, m_N$ defined by $m_i(S) = N\int_S p_i dm_0$; *the class of standard experiments is essentially equivalent to the class of probability measures over P with mean* $(1/N, \ldots, 1/N)$. The m_0 of the standard experiment of an experiment α will be called the *standard measure* of α; for two standard measures M, m of experiments α, β, the notation $M \supset m$ means that $\alpha \supset \beta$.

The following theorems, proved in [2], are valuable tools in the actual comparison of two experiments.

THEOREM 4. *For two standard measures* M, m, $M \supset m$ *if and only if for every continuous convex* $g(p)$, $\int g(p)\, dM \geqq \int g(p)dm$.

PROOF. Let A be the convex set determined by a finite set $a_i = (a_{i1}, \ldots, a_{iN})$, $i = 1, \ldots, k$, and define $L_i(p) = \sum_{j=1}^{N} a_{ij}p_j$, $L(p) = \min_i L_i(p)$, $f(p) = a_i$ when

$L_j(p) > L(p)$, $k < i$, $L_i(p) = L(p)$. Then $f \in (\alpha, A)$ for any standard experiment α, and for any $f^* \in (\alpha, A)$, $\sum_{j=1}^{N} p_j a_j^*(p) \geqq \sum_{j=1}^{N} p_j a_j(p)$ for all p, so that with $v^* = v(f^*)$, $v = v(f)$

$$\sum v_i^* = N \sum \int a_j^*(p)\, p_j dM \geqq N \sum_{j=1}^{N} \int a_j(p)\, p_j dM$$
$$= \sum_{j=1}^{N} v_j = N \int L(p)\, dM\,,$$

that is, if α has standard measure M,

$$\min_{v \in (\alpha, A)} \sum v_j = N \int L(p)\, dM\,.$$

Thus for a pair of standard experiments α, β with standard measures M, m, condition (3) of theorem 2 holds for every A determined by a finite set if and only if $\int L(p) dM \leqq \int L(p) dm$ for every $L(p)$ which is the minimum of a finite number of linear functions, that is, if and only if $\int c(p) dM \geqq \int c(p) dm$ for every $c(p)$ which is the maximum of a finite number of linear functions. It is readily shown by approximation that if condition (3) of theorem 2 holds for every A determined by a finite set, it holds for all A, and that $\int c(p) dM \geqq \int c(p) dm$ for all $c(p)$ which are maxima of a finite number of linear functions implies the same inequality for all convex $c(p)$, and the theorem follows.

THEOREM 5. *If* $N = 2$, $M \supset m$ *if and only if* $\int_0^y F_M(x)\, dx \geqq \int_0^y F_m(x)\, dx$ *for all* y, *where* $F_M(x) = M\{p_1 \leqq x\}$, $F_m(x) = m\{p_1 \leqq x\}$.

PROOF. Define $c_y(x) = y - x$ for $x \leqq y$, $c_y(x) = 0$, $x \geqq y$. Every convex function $c(x)$ on (0, 1) can be uniformly approximated by a linear function plus functions of the form $\sum_{i=1}^{K} a_i c_{y_i}(x)$, where $a_i \geqq 0$, so that, from theorem 4, $M \supset m$ if and only if $\int c_y(x) dM \geqq \int c_y(x) dm$ for all y. Now $\int c_y(x)\, dM = \int_0^y (y - x)\, dM$ $= \int_0^y F_M(x)\, dx$, integrating by parts, and similarly for $\int c_y(x) dm$, so that the proof is complete.

5. Sufficiency

A standard experiment α with measure M is said to be *sufficient* for a standard experiment β with measure m, written $\alpha \succ \beta$ or $M \succ m$, if there is a function $Q(p, E)$, defined for each $p \in P$ and each Borel set E of P such that (1) for fixed p, Q is a probability measure over P, (2) for fixed E, Q is a Borel function of p, and (3) for every E, $m_i(E) = \int Q(p, E) dM_i(p)$, $i = 1, \ldots, N$, where $m_1, \ldots, M_N$, $M_1, \ldots, M_N$ are the measures over P associated with m, M respectively, that is, if there is an experiment γ over the space $P_1 \times P_2$ with measures m_i^* such that the distributions of p_1, p_2 with respect to m_i^* are M_i, m_i and that p_1 is a sufficient statistic for (p_1, p_2) with respect to $m_1^*, \ldots, m_N^*$. That the second formulation is equivalent to the first follows from an unpublished result of Doob that conditional distributions of real or vector variables with respect to real or vector variables can always be defined so as to be probability measures; we shall use this fact several times in what follows. Essentially, $M > m$ means that, if p is the result of experiment M, then a vector p' selected according to the distribution $Q(p, E)$ will be as informative as a p^* resulting from experiment m, in the sense that for each i, p' and p^* have the same distribution.

THEOREM 6. *$M > m$ if and only if there is a function $D(p, E)$ such that* (4) *for fixed p, D is a probability measure over P,* (5) *for fixed E, D is a Borel function of p,* (6) $\int p_i dD(p^*, p) = p_i^*$, *and* (7) *for every E,* $M(E) = \int D(p, E)dm(p)$.

PROOF. Suppose $M > m$, and let i, p_1, p_2 be chance variables whose joint distribution is specified as follows: $i = 1, \ldots, N$, each with probability $1/N$; the conditional distribution of p_1 given i is M_i; and the conditional distribution of p_2 given i, p_1 is $Q(p_1, E)$, a function of p_1 only. Then p_1, p_2 have distributions M, m respectively, and m_i is the conditional distribution of p_2 given i. There is a determination of $D(p_2, E)$, the conditional probability given p_2 that $p_1 \in E$, such that for each p_2, D is a probability measure over P, and for any $g(p_1)$, $E(g|p_2) = g(p)dD(p_2, p)$. This D then satisfies conditions (4), (5), and (7) of the theorem, and (6) will be proved if we show that $p_{2i_0} = E(p_{1i_0}|p_2)$ for $i_0 = 1, \ldots, N$, where p_{ki} is the i-th coordinate of p_k, $k = 1, 2$.

We first verify that the probability $Pr\{i = i_0|p_k\} = p_{ki_0}$. This is equivalent to the statement that, for any S, $Pr(i = i_0, p_1 \in S) = \int_S p_{i_0} dM$, and a similar statement with M replaced by m for $k = 2$. Since $N_{p_{i_0}}$ is the density of M_i with respect to M,

$$\int_S p_{i_0} dM = \frac{1}{N} M_i(S) = Pr\{i = i_0\} Pr\{p_1 \in S \mid i = i_0\},$$

and similarly for $k = 2$. Moreover, $Pr\{i = i_0|p_2\} = E\{Pr(i = i_0|p_1, p_2)|p_2\}$, so that to show that $p_{2i_0} = E(p_{1i_0}|p_2)$, it is sufficient to show that $E\{Pr(i = i_0|p_1, p_2)|p_2\} = E(p_{1i_0}|p_2)$, and this will follow from (8) $Pr\{i = i_0|p_1, p_2\} = Pr\{i = i_0|p_1\}$. We postpone the proof of (8).

Now suppose there is a function D satisfying the conditions of the theorem. Let i, p_1, p_2 be chance variables whose joint distribution is specified as follows: p_2 has distribution m; the conditional distribution of p_1 given p_2 is $D(p_2, E)$; and the conditional probability that $i = i_0$ given p_1, p_2 is p_{1i_0}, a function of p_1 only. Condition (6) says that $E(p_{1i}|p_2) = p_{2i}$, so that $Pr\{i = i_0|p_2\} = E\{Pr(i = i_0|p_1, p_2)|p_2\} = E(p_{1i}|p_2) = p_{2i}$, and condition (7) guarantees that p_1 has distribution M. We next show that $Pr\{p_1 \in E|i\} = M_i(E)$, $Pr\{p_2 \in E|i\} = m_i(E)$, that is, that $Pr\{i = i_0, p_1 \in E\} = Pr\{i = i_0\}M_i(E)$ and $Pr\{i = i_0, p_2 \in E\} = Pr\{i = i_0\}m_i(E)$. Since $Pr\{i = i_0|p_1\} = p_{1i_0}$, $Pr\{i = i_0, p_1 \in E\} = \int_E p_{i_0} dM = M_i(E)/N$; similarly, $Pr\{i = i_0, p_2 \in E\} = m_i(E)/N$, so that we need simply note that $Pr\{i = i_0\} = \int p_{i_0} dM = 1/N$, since M is a standard measure.

Let $Q(p, E)$ be the conditional distribution of p_2 given p_1. Then requirements (1), (2) hold. Requirement (3) may be written $Pr\{p_2 \in E|i\} = E\{Pr(p_2 \in E|p_1)|i\}$, or $E\{Pr(p_2 \in E|p_1, i)|i\} = E\{Pr(p_2 \in E|p_1)|i\}$ which will follow from (9) $Pr\{p_2 \in E|p_1, i\} = Pr\{p_2 \in E|p_1\}$.

The proof of the theorem is now complete except for (8) and (9), which are special cases of

THEOREM 7. *If x, y, z are chance variables such that the distribution of z given x, y is a function of y only, then the distribution of x given y, z is a function of y only.*

PROOF. If $h(y, z)$ is the characteristic function of a set depending only on y, z and $g(x)$ is the characteristic function of a set depending only on x, we must show that $E(gh) = E[E(g|y)h]$. We prove the equation when $h(y, z) = h_1(y)h_2(z)$; the general result follows by approximation. We have $E[E(g|y)h_1h_2] = E\{E[gh_1E(h_2|y)]|y\} = E[gh_1E(h_2|y)] = E[gh_1E(h_2|x, y)] = E(gh_1h_2)$. This completes the proof.

Theorem 7 asserts essentially that a Markoff chain is also a Markoff chain in reverse, a fact noted in varying degrees of generality by several writers. The proof given here seems particularly simple.

Theorem 6 can be restated as follows: $M > m$ if and only if there are chance variables p_1, p_2 with distributions M, m such that $E(p_1|p_2) = p_2$.

THEOREM 8. *If $M > m$, then $M \supset m$.*

PROOF. For every continuous convex $g(p)$, $\int g(p)dM = \int \left[\int g(p)dD(p', p)\right] dm(p')$, where D is the set of measures whose existence is asserted by theorem 6. Since g is convex, $\int g(p)dD(p', p) \geqq g\left[\int pdD(p', p)\right] = g(p')$, so that $\int g(p)dM \geqq \int g(p)dm$ and $M \supset m$.

Thus theorems 4 and 6 reduce theorem 8 to a special case of the fact, noted by Hodges and Lehmann [4] and Doob (unpublished manuscript) that for any continuous convex g and any chance variables x, y, $E[g(x)] \geqq E\{g[E(x)|y]\}$.

6. Equivalence of $>$ and $\supset$ for $N = 2$

In this section we consider only the case $N = 2$, so that $P = \{(p_1, p_2)\}$, $p_i \geqq 0$, $p_1 + p_2 = 1$. For simplicity of notation, we denote the point (p_1, p_2) by the number $x = p_1$, $0 \leqq x \leqq 1$, so that a standard measure becomes simply a probability measure defined for Borel subsets of (0, 1) such that $\int_0^1 xdM = \frac{1}{2}$. For any standard measure M, we write $F_M(y) = M\{x \leqq y\}$, $c_M(y) = \int_0^y F_M(x)\,dx$. Then c_M is a nondecreasing convex function of y, $c_M(0) = 0$, $c_M(1) = \frac{1}{2}$, and, according to theorem 5, $M \supset m$ if and only if $c_M(y) \geqq c_m(y)$ for all y.

A class of measures $D(x, E)$ such that D is for each $x \in (0, 1)$ a probability measure over (0, 1), for each E a Borel function of x, and $\int_0^1 y\,dD(x, y) = x$ is called a transformation T, and for any standard measure m, the standard measure $M(E) = \int D(x, E)dm$ will be denoted by Tm. Theorem 6, for $N = 2$, asserts that $M \supset m$ if and only if there is a transformation T with $Tm = M$.

THEOREM 9. *For any sequence of transformations* $T_1, T_2, \ldots$, *there is a transformation* T *such that for any standard measure* m, $F_{m_k}(y) \rightarrow F_{Tm}(y)$ *at every point of continuity of* F_{Tm}, *where* $m_k = T_k \ldots T_1 m$.

PROOF. Let Ω be the space of sequences $\omega = (x_0, x_1 \ldots)$, $0 \leqq x_i \leqq 1$. For any a, $0 \leqq a \leqq 1$, there is a probability measure P_a, defined for Borel sets of Ω, such that $P_a\{x_0 = a\} = 1$ and $P_a\{(x_k \in E | x_0, \ldots, x_{k-1})\} = D_k(x_{k-1}, E)$, where D_k is the set of measures defining T_k. Then $E(x_{k+1} | x_0, \ldots, x_k) = x_k$, so that, by induction on j, $E(x_{k+j} | x_0, \ldots, x_k) = E[E(x_{k+j} | x_0, \ldots, x_{k+j-1}) | x_0, \ldots, x_k] = E(x_{k+j-1} | x_0, \ldots, x_k) = x_k$ for all $j \geqq 1$. Thus, $x_0, x_1, \ldots$ is a martingale; since $0 \leqq x_k \leqq 1$, a theorem of Doob [3] asserts that there is a chance variable x such that $x_k \rightarrow x$ with probability 1, and that $E(x | x_0, \ldots, x_k) = x_k$. In particular $E(x) = E(x_0) = a$. Let $D(a, E) = P_a\{x \in E\}$. We shall show that the set of measures $D(a, E)$, $0 \leqq a \leqq 1$, is the required transformation T.

For any Borel function $g(x_0, \ldots, x_k)$ (10) $\int g dP_a = \int\int \ldots \int g(x_0, \ldots, x_k) dD_k(x_{k-1}, x_k) \ldots dD_1(x_0, x_1) dI_a(x_0)$, where I_a is the measure concentrated at a, so that $\int g dP_a$ is a Borel function of a. The class $\mathcal{S}$ of sets S for which $P_a(S)$ is a Borel function of a is a normal class [7, p. 83] which includes all $(x_0, \ldots, x_k)$-Borel sets, so that $\mathcal{S}$ [5, p. 83] includes all Borel sets of Ω. In particular, $P_a\{x \in E\} = D(a, E)$ is a Borel function of a, so that $D(a, E)$ is a transformation T. For any standard measure m, define, for all Borel subsets S of Ω, $P_m(S) = \int P_a(S) dm(a)$. Then for every $g(\omega)$, $\int g dP_m = \int \left\{ \int g dP_a \right\} dm(a)$. Letting g be the characteristic function of an x_k-set and using (10) shows that the distribution of x_k is m_k. Also the distribution of x is Tm, and $x_k \rightarrow x$ with P_m-probability 1, so that $F_{m_k}(y) \rightarrow F_{Tm}(y)$ at all points of continuity of F_{Tm}.

THEOREM 10. *For* $N = 2$, *if* $M \supset m$, *then* $M > m$.

PROOF. We shall construct a sequence of transformations $T_1, T_2, \ldots$ such that $c_{m_k}(y) \rightarrow c_M(y)$ for all y, where $m_k = T_k \ldots T_1 m$. Then $c_M(y) = c_{Tm}(y)$ for all y, where T is the transformation whose existence is asserted in theorem 9, so that $M = Tm$. For any subinterval (a, b) of $(0, 1)$, let $T(a, b)$ be the transformation defined by

$$D(x, E) = \frac{b-x}{b-a} I_a + \frac{x-a}{b-a} I_b \qquad \text{for } a \leqq x \leqq b,$$

$$D(x, E) = I_x \qquad \text{for } x \text{ outside } (a, b).$$

It is easily verified that for any measure m, $c_{T(a,b)m} = c_m$ for x outside (a, b),

$$c_{T(a,b)m} = \frac{b-x}{b-a} c_m(a) + \frac{x-a}{b-a} c_m(b) \qquad \text{for } a \leqq x \leqq b.$$

Since $M \supset m$, $c_M(x) \geqq c_m(x)$ for all x. At any point $[t_1, c_m(t_1)]$ of the curve $y = c_M(x)$, draw a tangent, intersecting $y = c_m(x)$ say at $x = a_1$, $x = b$, where $a_1 \leqq t_1 \leqq b_1$. Then, with $T_1 = T(a_1, b_1)$, $c_{T_1 m} \leqq c_M$ with equality at $x = t_1$. Applying the same process to $y = c_{T_1 m}$ from a point $[t_2, c_M(t_2)]$ and continuing in this way, using a sequence $t_1, t_2, \ldots$ dense in $(0, 1)$, yields a sequence $T_1, T_2, \ldots$ such that $c_{m_k}(y) \rightarrow c_M(y)$ for all y.

Theorems 6 and 10 combine to yield the following partial converse of the result of Hodges and Lehmann and Doob mentioned in section 5: *If M, m are standard measures in* (0, 1) *such that* $\int g(x)dM \geqq \int g(x)dm$ *for every continuous convex* g, *then there are chance variables* p_1, p_2 *with distributions* M, m *such that* $E(p_1 | p_2) = p_2$. The requirement that M, m be standard measures on (0, 1) can be immediately weakened so that M, m can be any probability measures over a bounded interval (a, b). The extension to probability measures over $(-\infty, \infty)$ has not been carried out, and the extension to N-dimensional vector variables which, in view of theorem 6, would imply the equivalence of $\succ$ and $\supset$, remains unsolved. It has been pointed out by S. Sherman that theorems 5, 8, and 10, for the special case of measures concentrated at a finite number of points, are given, somewhat disguised, in [5, theorem 45 and associated results].

7. Combinations of experiments

For two experiments α, β, the *combination* (α, β) is the experiment defined by the space $X \times Y$ with the N probability measures $u_1 \times v_1, \ldots, u_N \times v_N$, where $\alpha = (X, u_1, \ldots, u_N)$, $\beta = (Y, v_1, \ldots, v_N)$.

THEOREM 11. *If* α^*, β^* *are the standard experiments for* α, β, *then the standard experiment for* (α^*, β^*) *is the same as that for* (α, β).

PROOF. If $Np_i(x)$, $Nq_i(y)$ are the densities of u_i, v_i with respect to u_0, v_0, then $d_i(x, y) = Np_i(x) q_i(y) / \sum_i p_i(x) q_i(y)$ is the density of $u_i \times v_i$ with respect to $w_0 = N^{-1} \sum_i u_i \times v_i$. The measure m for the standard experiment for (α, β) is the joint distribution of $d_1, \ldots, d_N$ with respect to w_0. The function $D_i(p, q) = Np_i q_i / \sum_i p_i q_i$ is the density for the measure $m_i \times M_i$ on $P \times Q$ with respect to the measure $\gamma_0 = N^{-1} \sum_i m_i \times M_i$, where $a^* = (P_1, m_1, \ldots, m_N)$, $\beta^* = (Q, M_1, \ldots, M_N)$, and the measure M for the standard experiment for (α^*, β^*) is the joint distribution of $D_1, \ldots, D_N$ with respect to γ_0. Now for each i, p has the same distribution with respect to m_i as $p(x)$ with respect to u_i, and similarly for q, M_i, $q(y)$, v_i, so that (p, q) with respect to $m_i \times M_i$ has the same distribution as $[p(x), q(y)]$, with respect to $u_i \times v_i$. Since D_i is the same function of p, q that d_i is of $p(x)$, $q(y)$, the joint distribution of $d_1, \ldots, d_N$ with respect to w_0 is the same as that of $D_1, \ldots, D_N$ with respect to γ_0.

THEOREM 12. *If* $\alpha_1 \succ \alpha_2$ *and* $\beta_1 \succ \beta_2$ *then* $(\alpha_1, \beta_1) \succ (\alpha_2, \beta_2)$.

PROOF. Since $\succ$ is transitive (this follows from theorem 6), we may suppose that $\alpha_1 = \alpha_2 = \alpha$; the general result would follow from this case, since $(\alpha_1, \beta_1) \succ (\alpha_1, \beta_2) \succ (\beta_1, \beta_2)$. Let α, β_1, β_2 have standard measures m, m', m'' and let $X = P_1 \times P_2 \times P_3 \times P_4$; we define a measure w_i on X by the following specifications: (p_1, p_2) have distribution $m_i \times m_i'$, and the conditional distribution of (p_3, p_4) for fixed p_1, p_2 is given by $Pr\{p_3 \in S, p_4 \in T | p_1, p_2\} = g(p_1)Q(p_2, T)$, where g is the characteristic function of S and Q is the function whose existence is implied by $\beta_1 \succ \beta_2$, so that $m_i''(T) = \int Q(p, T)dm_i'$. Then (p_3, p_4) have distribution $m_i \times m_i''$ with respect to w_i. The standard experiments for (α, β_1), (α, β_2) have measures $(M_1, \ldots, M_N)$, $(M_1^*, \ldots, M_N^*)$, where M_i, M_i^* are the distributions of $d = (d_1, \ldots, d_N)$ and $D = (D_1, \ldots, D_N)$ with respect to w_i, where $d_i = p_{1i}, p_{2i} / \sum_i p_{1i}, p_{2i}$ and $D_i = p_{3i}, p_{4i} / \sum_i p_{3i}, p_{4i}$, and it is sufficient to show that the conditional distribution of D given d is independent of i. For any function $f(D)$, in fact for any function of (p_3, p_4), $E(f \mid p_1, p_2) = h(p_1, p_2)$ is independent of i, so that we need show only that $E(h \mid d)$ using measure $m_i \times m_i'$ on $P_1 \times P_2$ is independent of i. Since the density of $m_i \times m_i'$ with respect to $\frac{1}{N} \sum_i m_i \times m_i'$ is d_i, a function of d, we conclude by Neyman factorization [4], that d is a sufficient statistic for the N measures $m_i \times m_i'$, so that $E(h \mid d)$ is independent of i.

The extension of the concept of combination of two independent experiments and of theorem 12 to the case of combination of n independent experiments is straightforward, and we obtain that if $\alpha_i \succ \beta_i$, $i = 1, \ldots, n$ then $(\alpha_1, \ldots, \alpha_n) \succ (\beta_1, \ldots, \beta_n)$. In particular if $\alpha \succ \beta$, then the experiment yielding n independent α's is sufficient for the experiment yielding n independent β's. It would be interesting to know whether conversely $(\alpha, \alpha) \succ (\beta, \beta)$ implies $\alpha \succ \beta$.

8. Binomial experiments

If the space X consists of two points, say 0, 1, an experiment α is simply the specification of a vector $a = (a_1, \ldots, a_N)$, $0 \leqq a_i \leqq 1$, where $a_i = m_i\{x = 1\}$. For the case $N = 2$, a simple computation shows that the standard measure M for (a_1, a_2) assigns measures d, $1 - d$ to the points $(p_1, 1 - p_1)$, $(p_2, 1 - p_2)$, where $d = (a_1 + a_2)/2$, $p_1 = a_1/2d$, $p_2 = (1 - a_1)/2(1 - d)$. Thus if $a_1 \leqq a_2$, we have

$$c_m(x) \begin{cases} = 0 & \text{for } 0 \leqq x \leqq p_1 \\ = d\,(x - p_1) & \text{for } p_1 \leqq x \leqq p_2 \\ = d\,(p_2 - p_1) + (x - p_2) & \text{for } p_2 \leqq x \leqq 1\ ; \end{cases}$$

if $a_2 \leqq a_1$, we interchange a_1, a_2 and replace d by $1 - d$ in the above description. For two binomial experiments $(a_1, a_2) = a$, $(b_1, b_2) = b$ with standard measures M, m, the relation between c_M and c_m is geometrically clear:

$$a \succ b \qquad \text{if and only if}$$

$\min\,[p_1(a), p_2(a)] \leqq \min\,[p_1(b), p_2(b)]$ and $\max\,[p_1(a), p_2(a)] \geqq \max\,[p_1(b), p_2(b)]$.

As an application of the comparison of binomial experiments, we consider the following 2×2 table problem. There are two characteristics H, S, whose proportions h, s, in the general population are known. Moreover it is known that the proportion of HS in the general population is either hs or a definite alternative c. A sample of size k is to be selected, after which some action is to be taken, whose worth depends only on whether $Pr\{HS\} = hs$ or $Pr\{HS\} = c$. Suppose that, for each observation, the statistician may select an individual at random from H or S or non-H or non-S; he has a choice among four binomial experiments which we denote by α_H, α_S, α_{CH}, α_{CS}. If it should happen that one of these, say α_H, is more informative than each of the other three, then it follows from the extension of theorem 12 that a sample of k individuals from H is more informative than any other combination of k experiments from α_H, α_S, α_{CH}, α_{CS} (a sample of k individuals from H can then also be shown to be more informative than any other *sequentially* selected set of k experiments from α_H, α_S, α_{CH}, α_{CS}, where the decision about which of the four experiments to do next depends on the results already obtained, but we shall not go into this).

The four experiments are $\alpha_H = (s, c/h)$, $\alpha_S = (h, c/s)$, $\alpha_{CH} = [s, (s-c)/(1-h)]$, and $\alpha_{CS} = [h, (h-c)/(1-s)]$. Computation of p_1, p_2 for each of the four experiments and using the condition given above for $a \succ b$ yields the following conditions:

For

$$
\begin{aligned}
H &\succ S &&: h \leqq s\\
H &\succ CH &&: h \leqq s\,,\ h+s \leqq 1\\
H &\succ CS &&: h+s \leqq 1\\
S &\succ CS &&: s \leqq h\,,\ s+h \leqq 1\\
S &\succ CH &&: h+s \leqq 1\\
CS &\succ CH &&: h \leqq s
\end{aligned}
$$

Without loss of generality, we may suppose that h is the smallest of the four numbers h, s, $1-h$, $1-s$. Then $\alpha_H \succ \alpha_S \succ \alpha_{CH}$, $\alpha_H \succ \alpha_{CS} \succ \alpha_{CH}$ and α_S, α_{CS} are not comparable unless $h = s$ or $h = 1 - s$. Thus *the procedure which always selects the characteristic which is rarest in the general population is more informative than any other procedure of the class considered.* The experiment α_{CH} is the least informative of the four, while α_S, α_{CS} are intermediate.

A second example, which suggests that for $N > 2$, the concept $\supset$ is quite strong (and $\succ$ is at least as strong as $\supset$), is the binomial experiment $(0, \frac{1}{2}, 1) = \alpha$. The standard measure M for α assigns measure $\frac{1}{2}$ to each of $Q_1 = (0, \frac{1}{3}, \frac{2}{3})$ and $Q_2 = (\frac{2}{3}, \frac{1}{3}, 0)$. Theorem 4 shows that the measures $m \subset M$ are exactly those concentrated on the line segment joining Q_1, Q_2; the binomial experiments $\beta = (a_1, a_2, a_3)$ whose m is concentrated on this line are those for which $a_2 = (a_1 + a_3)/2$. Thus α is not more informative than $(0, \frac{1}{2}, \frac{1}{2})$ or than $(\frac{1}{2} - \epsilon, \frac{1}{2}, \frac{1}{2} + 2\epsilon)$, $\epsilon > 0$ for instance, and for any $\beta \subset \alpha$, a suitable arbitrarily small perturbation of the a's destroys the relationship.

REFERENCES

[1] DAVID BLACKWELL, "The range of certain vector integrals," submitted to the *Proc. Amer. Math. Soc.*

[2] H. F. BOHNENBLUST, L. S. SHAPLEY and S. SHERMAN, private communication.

[3] J. L. DOOB, "Regularity properties of certain families of chance variables," *Trans. Amer. Math. Soc.*, Vol. 47 (1940), pp. 455–486.

[4] P. R. HALMOS and L. J. SAVAGE, "Applications of the Radon-Nikodym theorem to the theory of sufficient statistics," *Annals of Math. Stat.*, Vol. 20 (1949), pp. 225–241.

[5] G. H. HARDY, J. E. LITTLEWOOD and G. POLYA, *Inequalities*, Cambridge University Press, Cambridge, 1934.

[6] J. L. HODGES, JR. and E. L. LEHMANN, "Some problems in minimax point estimation," *Annals of Math. Stat.*, Vol. 21 (1950), pp. 182–197.

[7] S. SAKS, *Theory of the integral*, Hafner, New York, 1947.